

# Real Time Simulation of Power Grid Disruptions

## Project Description

DOE-OE and DOE-SC workshops (Reference 1-3) identified the key power grid problem that requires insight addressable by the next generation of exascale computing is coupling of real-time data streams (1-2 TB per hour) as the streams are ingested to dynamic models. These models would then identify predicted disruptions in time (2-4 seconds) to trigger the smart grid's self healing functions. This project attempted to establish the feasibility of this approach and defined the scientific issues, and demonstrated example solutions to important smart grid simulation problems. These objectives were accomplished by 1) using the existing frequency recorders on the national grid to establish a representative and scalable real-time data stream; 2) invoking ORNL signature identification algorithms; 3) modeling dynamically a representative region of the Eastern interconnect using an institutional cluster, measuring the scalability and computational benchmarks for a national capability; and 4) constructing a prototype simulation for the system's concept of smart grid deployment. The delivered ORNL enduring capability included: 1) data processing and simulation metrics to design a national capability justifying exascale applications; 2) Software and intellectual property built around the example solutions; 3) demonstrated dynamic models to design few second self-healing.

## Significance

The U.S. electric power system comprises multiple distinct interconnections of generators, high voltage transmission systems, and local distribution systems that maintain a continuous balance between generation and load with high levels of efficiency and reliability. This critical infrastructure has served the nation remarkably well, but is likely to see more changes over the next decade than it has seen over the past century. In particular, the widespread deployment of renewable generation, smart-grid controls, energy storage, and new conducting materials will require fundamental changes in grid planning and the way the power grid operates. Multiple distinct interconnections could be re-drawn or weakened and components separated widely geographically will be closely coupled. Rapid re-configuration in response to disruption signals will be likely over the next eight years.

## *Task 1- Simulate Real-Time Data Feeds – Speed, Quality, and Volume*

### Phase I. Research Aim -Simulate Real Time Data Feeds – Speed, Quality, and Volume

The objective of the phase 1 research was to develop and test a database system to receive the real-time data from frequency recorders system as a simulant for real-time synchro-phasor data. Currently in excess of 50 frequency recording units are porting real-time data to the ORNL. Critical to the success of Phase II was the uniquely documented data set collected in July 2009 (Cook NPP trip) that illustrated the coupled nature of the Eastern Interconnect within the frequency data. Within this data set, we have a unique, archived data to illustrate the expanded, coupled nature of smart grid control. This stream provided a real-time test stream where the data structure, speed, accuracy, latency, and signature potential can be assessed on live nation-wide data.

This test data set is unique and captured wide area data for significant grid disruption events. It is expected that for a nation-wide capability between 1 and 2 TB of data will be collected per hour and at least one year of data should be archived for both research and forensics. We developed the database structure necessary to deliver this data automatically to the signature identification algorithms while still trapping data errors, inconsistencies and data gaps. The structure and format of this database were important elements for the following reasons.

The total database size might reach 10 petabytes for active interrogation. The first inclination is to parse the data and solvers to separate processors. However, if the cause of the disruption is remote and creates anomalies throughout the grid, then this parallelization would mask the signature extraction. The database must be structured to handle this coupling. The deliverable for this phase was a prototype database and key data structure necessary to interface with in-house signature identification modules and dynamic models. We anticipate this data set will be useful as a standalone dataset research asset for a wide range of power grid investigations. The deliverable was that data set secured as intellectual property under copyright. This dataset took two months of real FNET datasets for June and July 2009 from several FDR devices mostly within the Eastern Interconnection. We then duplicated the given data sets and added white noise or a known signal as explained in the following

- Let  $X_1$  and  $X_2$  represent the actual data sets for June and July 2009 respectively;
- Replicate  $X_1$  and  $X_2$ ,  $N$  number of times each, where  $N$  is the limit when the size of the datasets is at least 0.25 TB; we generate four sets of replication, each of size at least 0.25TB; the total size of data generated equals at least 1TB.
- Let  $Y_1, Y_2, \dots, Y_N$  represent the first set of new data sets defined as:
- $Y_i = X_1 + SNR_i$ ,  $i = 1, \dots, N$ , where  $SNR$  in  $dB = -10, -8, -6, -4, \dots$
- $SNR$  (signal-to-noise ratio) =  $20 \log_{10} \frac{\text{norm} X_1}{\text{norm noise}}$ .
- Let  $Z_1, Z_2, \dots, Z_N$  represent the second set of new data sets defined as:
- $Z_i = X_2 + SNR_i$ ,  $i = 1, \dots, N$ , where  $SNR$  in  $dB = -10, -8, -6, -4, \dots$
- $SNR$  (signal-to-noise ratio) =  $20 \log_{10} \frac{\text{norm} X_2}{\text{norm noise}}$ .
- Let  $W_1, W_2, \dots, W_N$  represent the third set of new data sets defined as:
- $W_i = X_1 + P_i$ ,  $i = 1, \dots, N$ , where  $P$  is either a non-linear, stationary; linear, non-stationary; non-linear non-stationary signal. Different signals in each category are randomly selected from a library of signals.
- Let  $R_1, R_2, \dots, R_N$  represent the fourth set of new data sets defined as:  $R_i = X_2 + P_i$ ,  $i = 1, \dots, N$ , where  $P$  is either a non-linear, stationary; linear, non-stationary; non-linear non-stationary signal. Different signals in each category are randomly selected from a library of signals.  $R_i = X_2 + P_i$ ,  $i = 1, \dots, N$ , where  $P$  is either a non-linear, stationary; linear, non-stationary; non-linear non-stationary signal. Different signals in each category are randomly selected from a library of signals.

Using the generated data sets and the actual data sets, we summarized the expected performance of methods in Task 2 as follows:

1. The methods in Task 2 should detect the same events in (or perform the same way using)  $X_1$ ,  $Y_i$ , and  $W_i$  and the computation time should be about the same.
2. The methods in Task 2 should detect the same events in (or perform the same way using)  $X_2$ ,  $Z_i$ , and  $R_i$  and the computation time should be about the same.

***Task 2 – Rapid Extraction of Disruption Signatures – Develop and efficient algorithms for many core GPU architectures***

Using the database structure developed within Phase I, we interfaced the data set with fast signature recognition algorithms to identify the disruptions or impending failures within the grid. These signatures are necessarily more complex as more reactive elements are added to the grid. We will start with signatures based on the shape of the frequency variation and the time that the disruption is detected in multiple sensors. This disruption data then will be aggregated into a loss of component, node, or edge as input data to one or more national scale dynamic models that will then create a forecast or suite of forecasts within the 2-4 second decision loop.

As noted earlier, we required new algorithmic approaches as well as parallel formulations to address this new coupling. One of the critical components is the prediction of changes and detection of anomalies. The state-of-the-art algorithms are not suited to handle the demands of streaming data analysis. Recent DOE sponsored workshop on “Mathematics for Analysis of Petascale Data” have identified few challenges that are important for this proposal: (i) need for events detection algorithms that can scale with the size of data, (ii) need for algorithms that can not only handle multi-dimensional nature of the data, but also model both spatial and temporal dependencies in the data, which, for the most part, are highly nonlinear, (iii) need for algorithms that can operate in an online fashion with streaming data. New online anomaly detection techniques that take into account spatial, temporal, multi-dimensional aspects of the Phase I data set. The basic idea behind the approach is to (a) to convert a multi-dimensional sequence into a univariate time series that captures the changes between successive windows extracted from the original sequence using singular value decomposition (SVD), and then (b) to apply known anomaly detection techniques for univariate time series. A key challenge for the proposed approach is to make the algorithm scalable to big datasets by adopting techniques from perturbation theory for incremental SVD analysis. We used recent advances in tensor decomposition techniques which reduce computational complexity to monitor the change between successive windows and detect anomalies in the same manner as described above. For change detection we used Gaussian Process (GP) models which account for nonlinear dependencies in the data.

The rare event and GP based online change detection algorithms are computationally expensive. For example, hyper-parameter estimation of GP is  $O(n^3)$ . Therefore we propose to develop the parallel solutions on many core systems such as GPUs, because these algorithms involve lot of numerical operations and are highly data-parallelizable. The deliverable for this phase is a description the design requirements for the disruption identification

Obtaining real-time situational awareness presents a diverse set of computational requirements. These requirements can be categorized into three major goals. In real-time, process incoming data from distributed sensors (estimated at 2 TB/hour) and provide actionable intelligence to grid operators in a two to four second decision loop. This processing consists of two phases, processing sensor data and searching through the scenario library. The scenario library provides metrics including the approximate location of events, event type, and trip amount (MW loss).

We executed a large number of simulations to obtain frequency signatures for events representing each ( $n-k$ ) contingency. These signatures were collected, analyzed for significance, and stored in memory for fast searches while maintaining a high availability archive of one year of sensor data (estimated at 8.76PB) for batch analysis.

During the course of the study, an integrated prototype solution was developed and demonstrated using a mix of HPC technologies. Real-time event detection was accomplished through the development of the GAEDA (GPU-Accelerated Event Detection Algorithm) software. GAEDA leverages the high bandwidth of graphics processing units (GPUs), with each NVIDIA Tesla M2070 GPU capable of processing 1.2 GB/s of frequency data. The signature (scenario library) search was also accelerated using GPUs, attaining a rate of 1.5 million signatures per second per GPU.

These signatures were generated using the THYME simulation package for all ( $n-1$ ) contingencies in the Eastern interconnect, resulting in a total of 58,789 simulations. To support this large number of simulations, the team leveraged Keeneland, the NSF Track2D experimental HPC system. Keeneland consists of 120 nodes, each with a dual socket Intel Westmere CPU and an NVIDIA M2090 GPU. Analysis indicates that 1317/58789 cases (2.24%) exhibit frequency depression exceeding 8 mHz. Because 8 mHz represents the sensors limit of detection as established within Phase I, this represents a greater than 40 times faster searching capability when undetectable events are purged from the scenario library.

A design was also constructed for the archive of one year of sensor data. Utilizing compression, total data size is reduced from roughly 8.76PB to 7.01PB, a twenty percent reduction. Due to the batch nature of analysis on this data, a large-scale Hadoop cluster is ideal. At a minimum, this cluster would require an estimated 220 nodes, each with 16 2TB hard drives (which drive the overall cost of the system).

### Summary Table

Task	Requirement	Prototype Solution
Event Detection	Process 2TB/hr sensor data in real-time	GAEDA, 1.2 GB/s

<b>Signature Search</b>	Search all simulated scenarios in 2000ms	GAEDA, 1.5MM sig/s
<b>Scenario Library</b>	Exponential number of PG simulations	THYME on Keeneland, 58k simulations
<b>Sensor Data Archive</b>	Store 7.01 PB data	220 Node Hadoop Cluster

### ***Phase 3 Dynamic Modeling to Generate the Scenario Library***

One critical problem requiring high performance computing (HPC) is real-time power grid monitoring of large scale disruptions and look ahead forecast of future states. In this phase, we present a fast detection method for large scale power grid disruptions within the smart grid's 2-4 second decision loop. The method uses GPU accelerated algorithms to detect events based on an array of inexpensive frequency detectors that detect excursions resulting from a loss of one or more components such as generators, transmission lines, and transformers. A library of contingencies with the corresponding extracted signatures at the 50,000 buses of the Eastern Interconnection (EI) was generated with a 200 Teraflop computational platform that generated 50 million records to match against the measured signature.

We only considered the sensor-detectable records (contingencies) for the library instead of all possible scenarios to incredibly reduce the search and detection times. We used the criteria of being able to detect a 200MW disruption within 1000 msec, and leaving the rest of the time for transport of the measured frequency excursions (signatures). Our power flow analyses on the EI show that a 25MW disruption approximately generates about 1 mHz frequency excursion. Therefore, those records that show a frequency excursion of less than 8 mHz are considered as undetectable and thus eliminated. Fast searching of the library to present potential component losses and future states demonstrated the ability to search 3 million contingencies within the design requirement of 1000 msec. This places a practical limit of  $N-10$  contingencies for an EI sized grid.

To generate the contingency scenarios, time domain simulations are performed using ORNL's THYME simulator operating on ORNL's Keeneland cluster, which is a 200 Teraflop high performance computer. We present in the Summary Table the detected power grid components' statuses and their assessed alerts to the power system operators via an interactive overlay within a web-based Google Earth geographic platform. Numerical results are presented to demonstrate the accuracy and detection speed of the proposed approach.

### ***Task 4 Research Aim - Demonstration of Integrated Capability***

In Phase IV, we demonstrated the integrated capability during the large scale disruptions likely to occur on an approximately monthly basis within the national power grid. Based on the results in the previous three tasks, the prototype database, signature identification modules and dynamic forecasting modules will be prototyped into an integrated system. The objective under this task was the collection of metrics to scale this power grid application to illustrate whether these power grid problems rise to the level of an exascale computational design problem.

The simulation results from the THYME simulator were interfaced with different visualization frameworks (Google Earth platform and Tableau Software) to provide an understanding of the frequency variations over space and time for different contingency scenarios and useful insights into these datasets. Being able to port these big data results directly into commercial visualization tools like Tableau, STI (Space-Time-Insight) by displaying statuses/alerts and making them available directly on the commander's tablet and the control room has proved to be extremely helpful in presenting these results to decision makers. These visualizations display the results in a highly intuitive way by representing several parameter values with different visual attributes (color, size, height etc.) simultaneously.

**GridLab-D Feeder Model Simulation Results Visualization:** For GridLab-D, which is a power distribution system simulation and analysis tools, a VERDE(Google Earth platform) framework was developed to study the voltage fluctuations, effect of electric vehicle (EV) adoption on the distribution grid and to provide real time situational awareness due to weather events. Here one of the 24 generic Feeder models (R5\_12.47kv) which represents a moderately populated suburban area was geo-located over a Knoxville neighborhood and the simulations results were displayed as a time animation for a 12 hr period. Typically there are voltage drops at the end of the feeder line away from the substation creating an under-voltage condition. Such patterns can be easily identified in these heat map animations. These visualizations were developed to the power system operators in their control room almost real-time for decision-making process.

#### ***Task 5 Sensor Re-invention to support peta-scale applications***

A frequency measurement system design was developed to provide real-time measurements of frequency transients in an electric power system. The sensor has the characteristics of a) cyber-security using hardware-accelerated cryptography compliant to FIPS 140-2(<sup>1</sup>), b) producing a symbol rate of 8000 measurements per second, c) providing at that rate, average frequency with a sensitivity of 25mHz within 80ms, d) employing a timing system that can accurately time-stamp each measurement using VLF broadcast time-code signals and the IEEE 1588 precision time protocol standard, and e) reducing system cost to less than \$1000 per copy.

The sensor system prototype was based on the multi-core Digital Signal Processor (DSP) device. A block diagram design for the hardware components is developed, as well as software for the major sensor functions. A system simulation was performed on THYME-generated datasets to describe and verify the feature of c) above. The software files and all simulations are maintained in source code control. Critical hardware components to assemble the prototype have been obtained.

***Future Activities:*** Key to producing an enduring capability future funding is a Center for National grid simulation that will have a variety of DOE-OE missions. Once captured, this center will build on ORNL's long-standing expertise in running large-scale applications for the Department of Energy. The capture of a co-design center for the power grid strategically aligns with ORNL directions for both next generation computing and innovative analysis for DOE

---

<sup>1</sup> Validated to FIPS 140-1, pending re-certification to (current) FIPS 140-2

power grid programs. A dedicated institutional cluster runs programs solely for electric grid-related applications which include explicitly parallel planning, real-time dynamics visualization, run-time dynamic control predictions, and impacts analysis. The data for the visualization comes from real-time feeds from the utilities, OASIS web sites, and instrumented sensors that ORNL has deployed in the field.

### **Publications:**

M. Allen, "Preparing the Way for New Policy Regarding Adaptation of U.S. Electricity Infrastructure to Climate Change", ORNL/TM-2012/192, May 2012.

V. Chandola, Olufemi A. Omitaomu, and Steven J. Fernandez (2012). Data Analysis for Real Time Identification of Grid Disruptions. In Ting Yu, Nitesh Chawla, and Simeon Simoff (Eds.), *Computational Intelligent Data Analysis for Sustainable Development*. Taylor & Francis, London, UK.

S. Fernandez, "Microgrid Situational Awareness for DoD Installations" Advanced Microgrid Concepts and Technologies Workshop, Beltsville MD 7-8 June 2012.

J. Nutaro, "Split System Method for Simulating Cyber-Physical Systems", *Proceedings for Modeling, Simulation and Optimization for the 21<sup>st</sup> Century Electric Power Grid*. October 21-25, Lake Geneva, Wisconsin, USA.

M. Olama, K. Spafford, O. Omitaomu, S. Chinthavali, and S. Fernandez, "High Performance Computing for Real-Time Detection of Large Scale Power Grid Disruptions," *Proceedings of the Modeling, Simulation, and Optimization for the 21<sup>st</sup> Century Electric Power Grid Conference*, Oct. 21-25, 2012.

Olufemi A. Omitaomu, Kyle L. Spafford, and Steven J. Fernandez (2012). A GPU-based Real-time Event Detection Framework for Power System Frequency Data Streams, *Proceedings for Modeling, Simulation and Optimization for the 21<sup>st</sup> Century Electric Power Grid*. October 21-25, Lake Geneva, Wisconsin, USA.

B. Vacaliuc, J. J. Nutaro, D. B. Koch, B. E. Huey, and S. F. Smith, "A Frequency Data Recorder for Multiple Generator Tracking", *Modeling, Simulation, And Optimization for the 21st Century Electric Power Grid*, Lake Geneva, WI, Oct 2012, accepted for publication (PTS #35874)

B. Vacaliuc, J. J. Nutaro, D. B. Koch, B. E. Huey, and S. F. Smith, "A Low Cost Disciplined Sampling Clock using the WWVB (60KHz) Broadcast", *Collaboration for Astronomy Signal Processing and Electronics Research 2012 Annual Meeting*, Green Bank, WV, Aug 2012, oral presentation (PTS #37375)

B. Vacaliuc, J. J. Nutaro, B. K. Daniel, "Highly Accurate Frequency Tracking for Smart Grid Infrastructure", *Future of Instrumentation International Workshop 2012*, Gatlinburg, TN, October 2012, submitted for review (PTS #38013)

***Copyrighted Software***

UT-Battelle Copyright entitled “GPU Accelerated Event Detection Algorithm” (GAEDA).

***Patent Disclosures Filed***

UT-Battelle Invention Disclosure No. 2459 entitled “Real-Time Simulation of Power Grid Disruptions”

UT-Battelle Invention Disclosure No. 2554, entitled “The Verde Analytic Modules”.

B. Vacaliuc, Et. Al., “A Frequency Data Recorder with Multiple Generator Tracking,” DOE S-Number S-124,404; Invention disclosure 201202836, Mar 2012